Weekly Report

September 17, 2017

1 Work

- 本周完成了硕士论文开题报告的材料包括开题报告和文献综述。
- 验证了bhtsne（tsne作者对tsne的加速）和Largevis的效率比较，在最后优化的部分的速度还是bhtsne更快一些。对于MNIST数据集，bhtsne180秒，LargeVis300秒。
- 阅读了一些相关的降维方法的论文，如LLE。

1.1 目前认为可行的加速降维的方法（主要是加速优化部分的速度）

1. 提供一个更好的降维初始值。目前的算法（bhtsne，和Largevis）对于初始值都是随机分布的，提供一个好的初始值可以更快地看到降维结果。
    - 基于锚点，先投影一部分点下去，然后其余的点按照锚点的位置进行投影。
    - 基于点与点之间的相似度（概率）矩阵，用线形的方法如MDS先投影下去
    - 其他可以用于对kNN graph投影的方法
2. SNE系列的投影算法是最小化了两个概率分布p_{ij}和q_{ij}之间的距离。然而每一个的计算都需要$O(N^2)$。bhtsne利用kNN将p_{ij}降低到$O(kN)$，利用四叉树将q_{ij}降低到$O(Nlog(N))$。是否可以将kNN应用于q_{ij}的计算，进一步从$O(Nlog(N))$也降低到$O(kN)$。这方面有待验证，bhtsne的作者应该也尝试过吧。

3. CPU+GPU共同计算。我们可以将kNN和投影的算法同时进行计算，当计算出一个准确率比较低的kNN时，就可以用于投影的过程。利用CPU不断更新kNN结果，把最新的结果不断送入GPU，GPU根据最新的结果，继续迭代投影结果。这样的原因是由于《Approximated and User Steerable tSNE for Progressive Visual Analytics》中验证了即使是7%准确率的kNN，也能用于投影，得到还可以的投影结果。

1.2 工作进度

Table 1: 工作进度

任务	当前进度	截止时间
投影	正在进行对高维投影方法的讨论。	9月30日
*2Vec综述	收集与TransE相关的一些列文献	10月30日
专利	已经联系过律师，会尽快完成。	

2 Paper Reading

2.1 Beyond Tasks: An Activity Typology for Visual Analytics

作者基于Activity Theory在人机交互中的应用，提出了可视分析中的基本原则Activity Typology，主要包括Personas，Products，Capabilities，Contexts，Rules和Roles等六个元素。

2.2 BiDots: Visual Exploration of Weighted Biclusters

文章设计了一个针对Biclusters的可视分析系统，主要可以用来分析如文档、主题、单词等两两维度之间频繁出现的子集。 使用张量分解的方式从时序数据中提取主题并进行分析，

2.3 Translating Embeddings for Modeling Multi-relational Data

文章主要用于将知识图谱中的关系嵌入到高维空间。一个三元主(h, l, t)，其中h和l是对象，l是关系，我们最终希望能到$h + l = t$。所以优化的目标是三元主之间计算的误差要比其他随机对象的计算误差低。

Activity-Centered Design using the Activity Typology Model (Section 2)				
1. Define target activity		**2. Organize design targets**	**3. Identify design tensions**	**4. Generate target qualities**
Define the activity that is to be supported through the design and introduction of new tools. This can be a general activity that takes many different forms in practice. Use the fundamentals of Activity Theory (Section 1.2) to elaborate.		*Use each element of the activity typology model to organize related concepts from the many forms of the target activity. Each of these concepts is a potential target in the design of new tools.*	*Search for bipolar tensions that cut across the design targets of each activity element, e.g., identify attribute pairs relevant to all design targets but varying in their nature and extent.*	*Use each design tension as a springboard to generate creative design resolutions. Express these resolutions as the qualities of tool interactions that would resolve design tensions and thus support the many forms of the target activity in general.*
Activity Typology Model (2.1)		**Visual Analytics Activity Typology (Section 3)**		
Element	**Definition**	**Design Targets**	**Design Tension**	**Target Quality**
Personas	*types of people using the tools of the activity*	desk analyst, case investigator, field officer, domain expert, professional analyst	*Acting as:* Data collector *vs* Data analyst	**Portable Analysis** *ability to transfer analytic work across people, places, time, and devices*
Products	*types of outcome that motivate the activity*	derive insights, develop options, make arguments, present assessments, manage situations	*Acting to:* Make sense *vs* Make artifacts	**Presentable Analysis** *ability to curate presentable summaries of the analytic discovery process*
Capabilities	*types of task supported by the tools of the activity*	searching sources, visualizing data, reading reports, tagging interests, recording viewpoints	*Acting by:* Funneling data *vs* Testing data	**Perspectival Analysis** *ability to create and annotate workspaces with analytic perspectives*
Contexts	*types of contextual factor that shape the activity*	co-located teams, distributed teams, distributed communities, synchronicity, mobility	*Acting in:* Defined teams *vs* Defined areas	**Proxemic Analysis** *ability to locate oneself and others within the space of analytic views and viewpoints*
Rules	*types of constraint on the performance of activity*	relevance, confidence, provenance, access rights, time pressure	*Acting under:* Competing interpretations *vs* Competing demands	**Provisional Analysis** *ability to view and proactively reduce the uncertainty of analytic work at any time*
Roles	*types of coordinated contribution to the activity*	producers, consumers, responders, decision makers, policy makers	*Acting with:* Process partners *vs* Product partners	**Polymorphic Analysis** *ability to export analytic reports tailored by audience, format, and purpose*

Figure 1: Activity Typology

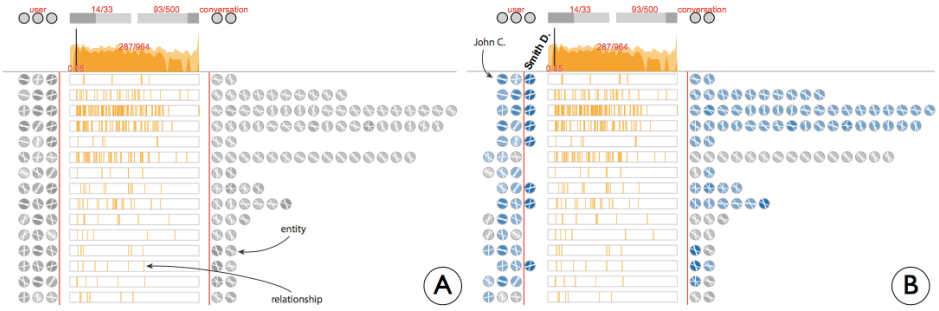

Figure 2: BiDots

2.4 Neighborhood Preserving Embedding(NPE)

线性降维的方法都是对原始的向量乘以一个转换矩阵$y = Ax$，求解转换矩阵的方法一般都是优化一个目标函数$\sum_i (y_i - \sum_j W_{ij} y_j)^2$。这篇文章目标函数的构造是基于最小化重构数据点的误差，数据点的重构是由它的最近邻线性组合构成。

2.5 Nonlinear Dimensionality Reduction by Locally Linear Embedding(LLE)

和上文求解的目标类似，然而并没有要求$y = Ax$，而是直接求解降维后的y值，因此可以说是非线形的。

2.6 Locality Preserving Projections(LPP)

基本求解方法与NPE类似，唯一不同的是对目标函数的构建，$\sum_i (y_i - y_j)^2 W_{ij}$，基本上意味着，越接近的数据点（最近邻）权重越大，因此降维之后要越接近。